gembloux agro bio tech

Université de Liège

University of Ljubljana, Slovenia
November 13th, 2012

Relationship matrices and Iterative construction of their inverses

Pierre Faux & Nicolas Gengler

University of Liège, Gembloux Agro-Bio Tech, Belgium

AFR Beneficiary (FNR)

Fonds National de la Recherche Luxembourg

WBI Wallonie-Bruxelles International

Summary

- Theoretical measurement of similarity between related
- Construction of the matrix **A**
- Direct creation of the inverse of **A**
- Iterative construction of inverses using the inverse factorization
- Application to the case of the inverse of **G**

Theoretical measurement of similarity between related individuals

(“Génétique Quantitative”, Gengler, 2012)

Relationships and Similarity

- When observing phenotypes:
  - A son has a certain similarity with his father
- Relationships might help while **predicting** unknown animals
- How **to scale this similarity**? Theoretical view:
  - What is similarity at a molecular level?
  - How to express this similarity from molecular to individual level?

Relationship Measurement

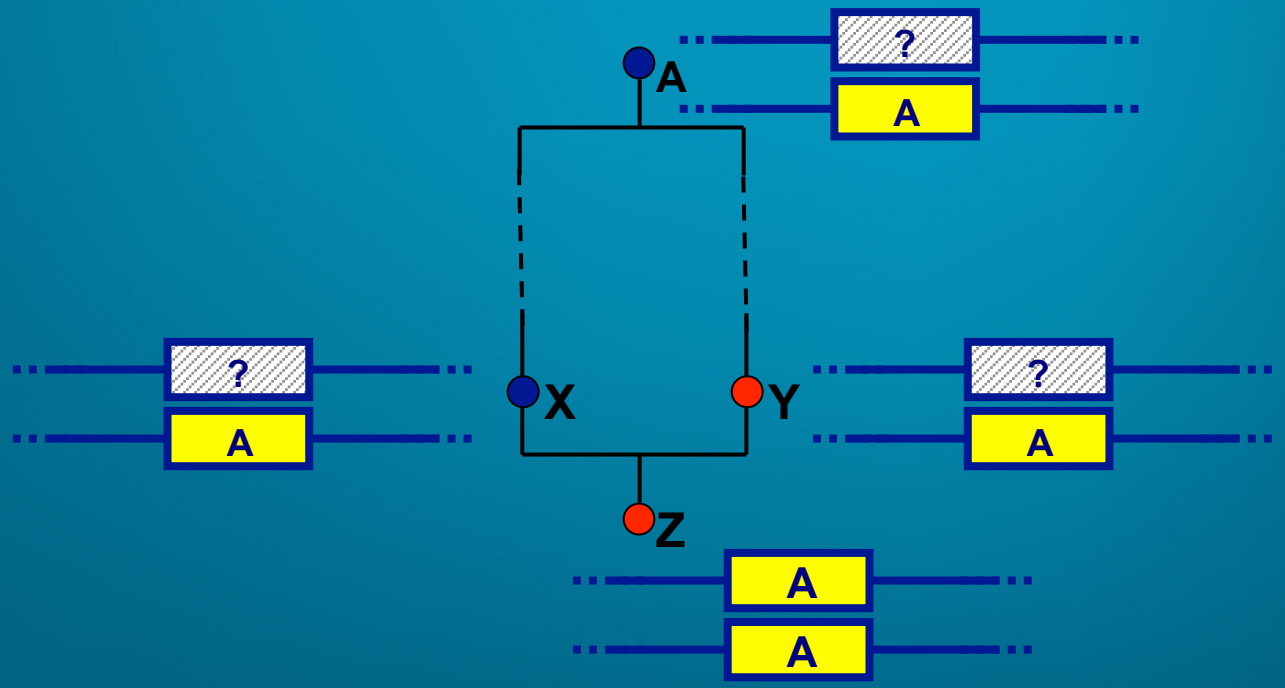
- Let us have a biallelic gene (either yellow or red)

- 2 possible configurations
  - Homozygous (identical)

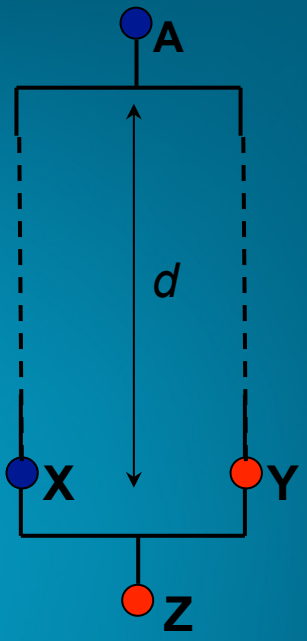
  - Heterozygous (non-identical)

Identical-By-Descent (IBD)

- Homozygosity is imputed to the same common ancestor
- The « allele's pathway » is known



Inbreeding

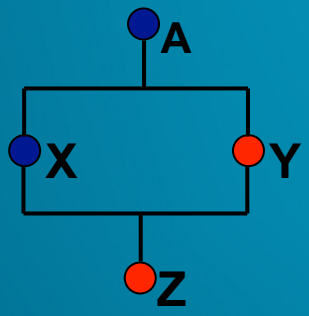


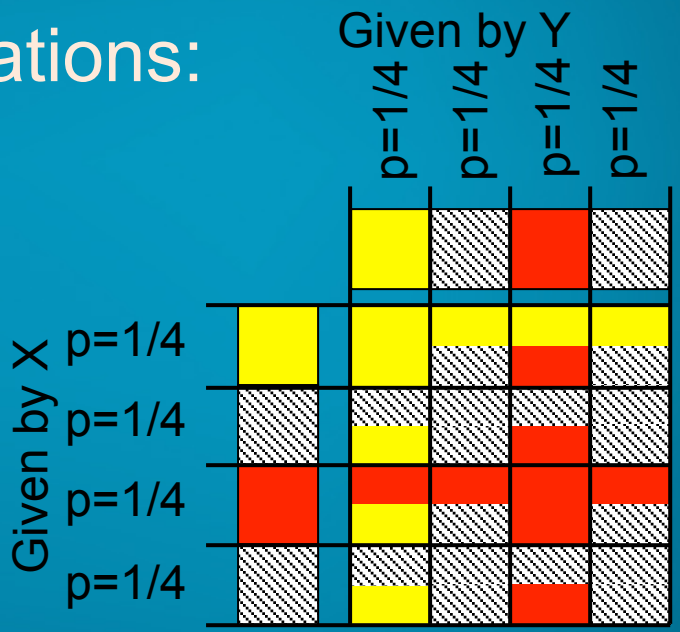
- IBD at a molecular level …

- … is expressed as inbreeding at the individual level
  - Inbreeding coefficient (***F***)= P(IBD), is function of *d*
  - Inbreeding coefficient = ½ of relationship coefficient ($a_{X,Y}$) between parents,
  - Relationship coefficient is only **additive part** of relationship

P(IBD)?

- How to assess P(IBD) for animal Z?
- Different gene configurations:



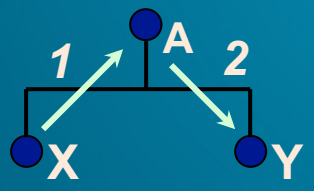


- Thus, P(IBD) = 2/16 = 0.125 = F_Z
- And $a_{X,Y}$ = 2 x 0,125 = 0.25

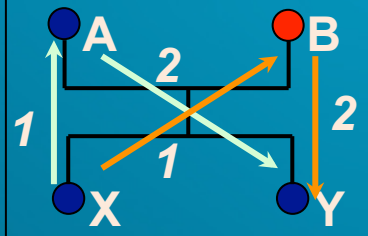
Easier method to assess P(IBD)

- Count ***n*** , lowest number of steps to join X to Y through each of ***x*** common ancestors
- Relationship coefficient between X and Y:

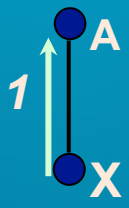
$$a_{X,Y} = (2^{-n})_1 + (2^{-n})_2 + \ldots + (2^{-n})_x$$



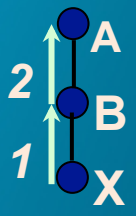
Halfsibs:
$2^{-2} = 0{,}25$



Sibs:
$2^{-2} + 2^{-2} = 0{,}5$



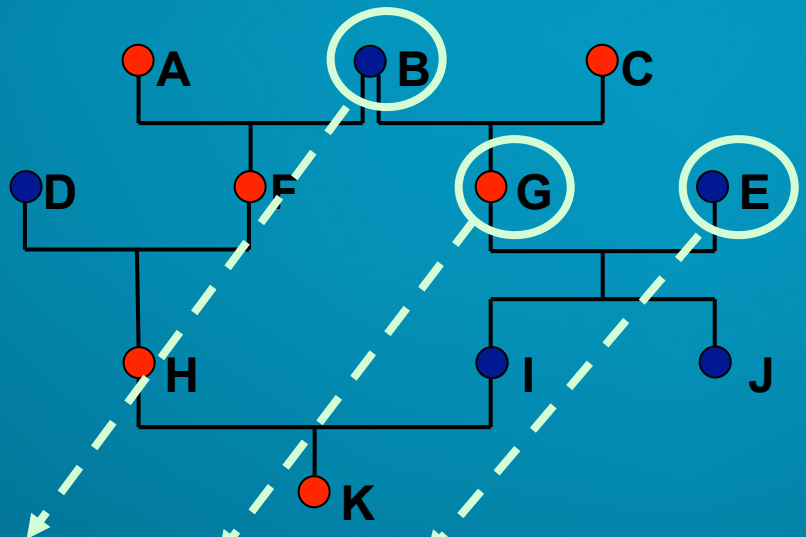
Child to parent:
$2^{-1}=0{,}5$



Grandchild to grandparent:
$2^{-2}=0{,}25$

Easier Method

- May ease more complicated cases
  - Example: various common ancestors at various levels



- $a_{K,J} = 2^{-5} + 2^{-3} + 2^{-3} = 0.28125$
  $= 1/32 + 1/8 + 1/8 = 9/32$

Construction of the additive genetic relationship matrix

Tabular Method: Views

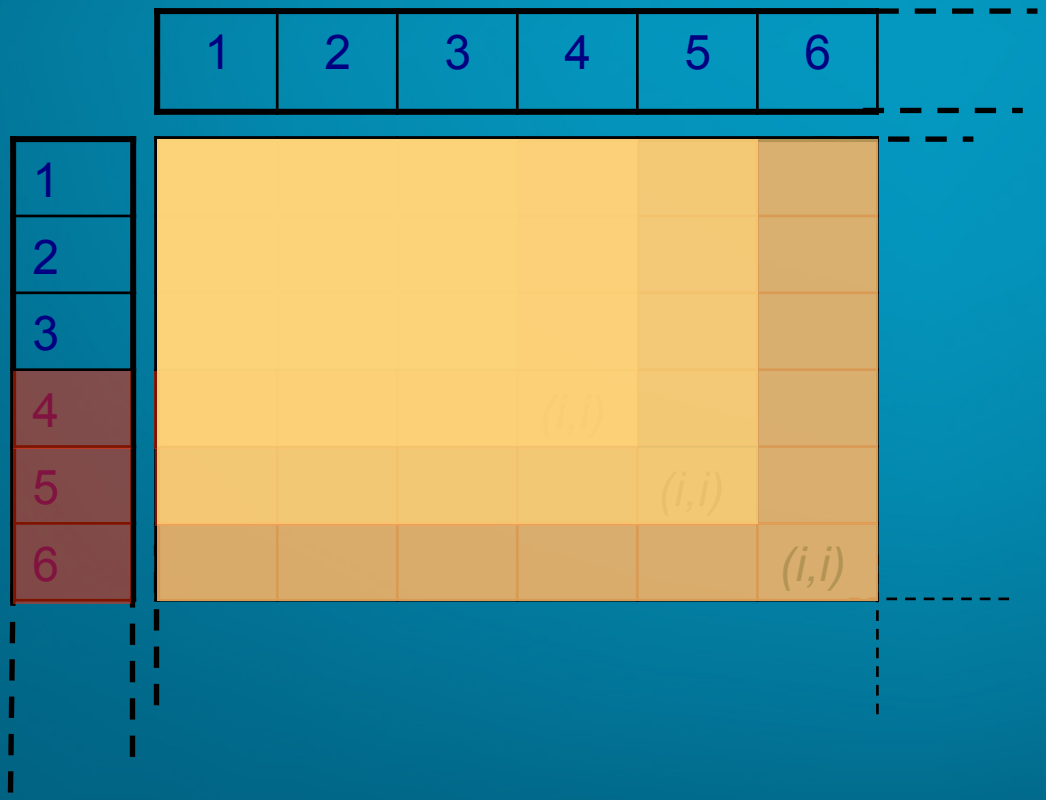
- The previous method needs to be streamlined
  - May be executed iteratively
  - Use of a symmetric matrix *n***n* (noted **A**) that will content relationships between all individual
- Requires to re-organize pedigree data
  - A given animal always needs to appear after both parent
- Proceeds by filling successive squares
  - From square 1*1 to square *n***n*
  - For animal *i*, firstly row *i* from 1 to *i*, including **element** (*i,i*), secondly **column** *i* from 1 to *i*-1 (using symmetry)
  - At each row, sum of **weighted** contributions from sire, dam, and animal himself

Tabular Method: Rules

1. For a given population of *n* individuals (sorted by generation order), define a square matrix of size *n*
2. At line *j*, for all *i* younger than *j*, the relationship between *i* and *j* is equal to the sum of the half of the relationship of known parents of *j* with *i*
3. At line *j*, diagonal element equal 1, plus half of the relationship between parents of *j*
4. At line *j*, paste line *j* (from 1 to *i*) in column *j*
5. Go back to point 2.

Tabular Method: Iterative process

- For animal *i*, firstly **row** *i* from 1 to *i*, including **element** (*i*,*i*), secondly **column** *i* from 1 to *i*-1



Tabular Method: Example

- Numbered pedigree :

Animal	Père	Mère
A	-	-
B	-	-
C	-	-
D	-	-
E	-	-
F	B	A
G	B	C
H	D	F
I	E	G
J	E	G
K	I	H
L	J	K

Animal	Père	Mère
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	2	1
7	2	3
8	4	6
9	5	7
10	5	7
11	9	8
12	10	11

Tabular Method: Example

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0	0	0	0	0,5						
2	0	1	0	0	0	0,5						
3	0	0	1	0	0	0						
4	0	0	0	1	0	0						
5	0	0	0	0	1	0						
6	0,5	0,5	0	0	0	1						
7												
8												
9												
10												
11												
12												

Tabular Method: Example

		1	2	3	4	5	6	7
0,5 *	2	0	1	0	0	0	0.5	*0*
+ 0,5 *	3	0	0	1	0	0	0	*0*
+ 1 *	7	*0*	*0*	*0*	*0*	*0*	*0*	1+0*1/2
=	7	0	0.5	0.5	0	0	0.25	1

Tabular Method: Example

	1	2	3	4	5	6	7	8	9	10	11	12
1	1					1/2		1/4			1/8	1/16
2		1				1/2	1/2	1/4	1/4	1/4	1/4	1/4
3			1				1/2		1/4	1/4	1/8	3/16
4				1				1/2			1/4	1/8
5					1				1/2	1/2	1/4	3/8
6	1/2	1/2				1	1/4	1/2	1/8	1/8	5/16	7/32
7		1/2	1/2			1/4	1	1/8	1/2	1/2	5/16	13/32
8	1/4	1/4		1/2		1/2	1/8	1	1/16	1/16	17/32	19/64
9		1/4	1/4		1/2	1/8	1/2	1/16	1	1/2	17/32	33/64
10		1/4	1/4		1/2	1/8	1/2	1/16	1/2	1	**9/32**	41/64
11	1/8	1/4	1/8	1/4	1/4	5/16	5/16	17/32	17/32	**9/32**	33/32	42/64
12	1/16	1/4	3/16	1/8	3/8	7/32	13/32	19/64	33/64	41/64	42/64	73/64

Direct creation of the inverse of the matrix A

Inverse of A

- Additive genetic relationship matrix **A** will structure the covariance of breeding values of animals in a related population

- Used in the LHS of the MME for an animal model:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda \cdot \mathbf{A}^{-1} \end{bmatrix}$$

Factorization of A

- Root-free Cholesky factorization of matrix **A**

$$\mathbf{A} = \mathbf{TDT}'$$

- Structures of matrices **T** and **D**:
  - **T**: for any animal *i*, with sire *s* and dam *d*, the row is filled in with the same rules as for **A** :

$$\mathbf{T}_i = 0.5\,\mathbf{T}_s + 0.5\,\mathbf{T}_d \quad ; \quad \mathbf{T}_{i,i} = 1$$

  - **D**: for any animal *i*, only one element, equal to

$$\mathbf{D}_{i,i} = \left\{1 \quad | \quad 0.75 - 0.25(F_p) \quad | \quad 0.5 - 0.25\,(F_s + F_d)\right\}$$

Inverse of the factorization of A

- Factorization of the inverse of **A**

$$\mathbf{A}^{-1} = (\mathbf{T}^{-1})' \, \mathbf{D}^{-1} \, \mathbf{T}^{-1}$$

- Structures of matrices $\mathbf{T}^{-1}$ and $\mathbf{D}^{-1}$:
  - $\mathbf{T}^{-1}$: for any animal *i*, with sire *s* and dam *d*, only 3 elements are filled in

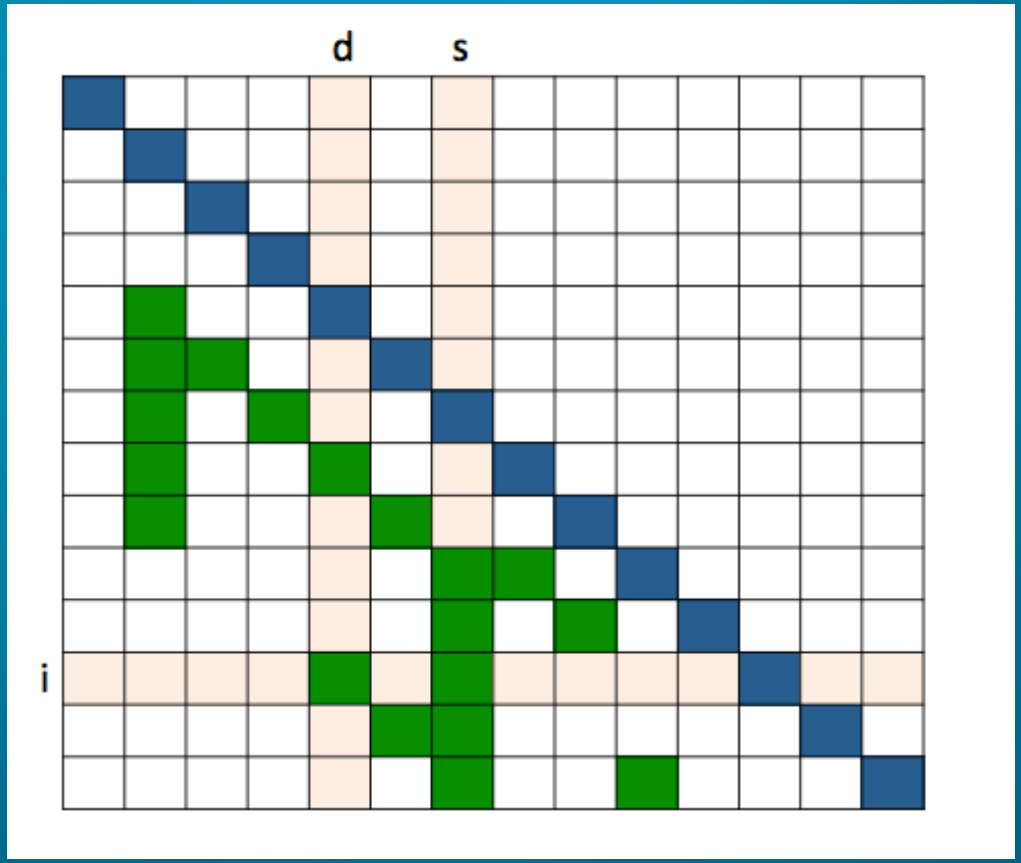
$$\mathbf{T}^{i,s} = -0.5 \quad ; \quad \mathbf{T}^{i,d} = -0.5 \quad ; \quad \mathbf{T}^{i,i} = 1$$

  - $\mathbf{D}^{-1}$: for any animal *i*, inverse of $\mathbf{D}_{i,i}$

Inverse of the factorization of A

- Structure of matrix $\mathbf{T}^{-1}$:

  d, s are parents of *i*



Direct creation of the inverse of A

- Inverse of **A** can be viewed as a sum of sparse matrix:

$$\mathbf{A}^{-1} = (\mathbf{A}^{-1})_1 + (\mathbf{A}^{-1})_2 + \ldots + (\mathbf{A}^{-1})_n$$

- Each sparse matrix is created as:

$$(\mathbf{T}^i)' \times \mathbf{D}^{i,i} \times \mathbf{T}^i$$

- Max. 3 non-zeros entries in $\mathbf{T}^i$, each sparse matrix is thus a block of max 9 non-zeros entries (with known positions: animal, sire, dam)
- Time for creation is thus linearly related to size of the population

Direct creation of the inverse of A

- Three possible configurations (when no inbreeding)
  - 0: | 1 | | 1 | | 1 | = | 1 |
  - 1: | -0,5 | | 4/3 | | -0,5 | 1 | = | 1/3 | -2/3 |
       | 1 | | | | | | | -2/3 | 4/3 |
  - 2: | -0,5 | | | | | | | 1/2 | 1/2 | -1 |
       | -0,5 | | 2 | | -0,5 | -0,5 | 1 | = | 1/2 | 1/2 | -1 |
       | 1 | | | | | | | -1 | -1 | 2 |

Iterative construction of inverses using the inverse factorization

Alternative way of direct creation

- Let us define:
  - $\mathbf{Z}_j$ = A, from row 1 to *i* and from column 1 to *i*
  - $\mathbf{y}_j$ = A, from row 1 to *i*, at column *j*
  - $\mathbf{b}_j = (\mathbf{Z}_j)^{-1}\,\mathbf{y}_j$
- For this particular case, computation of $\mathbf{b}_j$ is trivial
- Link with tabular method!

Alternative way of direct creation

- Let us now define $\mathbf{Z}_k$ and $(\mathbf{Z}_k)^{-1}$, (k=j+1)

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{Z}_j & \mathbf{y}_j \\ \mathbf{y}_j' & a_{jj} \end{bmatrix} \quad \Rightarrow \quad (\mathbf{Z}_k)^{-1} = \begin{bmatrix} (\mathbf{Z}_j)^{-1} + \alpha_j \mathbf{b}_j \mathbf{b}_j' & -\alpha_j \mathbf{b}_j \\ -\alpha_j \mathbf{b}_j' & \alpha_j \end{bmatrix}$$

where: $\alpha_j = (a_{jj} - \mathbf{b}_j' \mathbf{y}_j)^{-1}$

- Relation between $(\mathbf{Z}_k)^{-1}$ and $(\mathbf{Z}_k)^{-1}$

$$(\mathbf{Z}_k)^{-1} = \begin{bmatrix} (\mathbf{Z}_j)^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + \alpha_j \begin{bmatrix} -\mathbf{b}_j \\ 1 \end{bmatrix} \begin{bmatrix} -\mathbf{b}_j' & 1 \end{bmatrix}$$

Alternative way of direct creation

- As you can see ...

  $\mathbf{b}_j = - \mathbf{T}^{-1}$ (at row *j*, from column 1 to *i*)

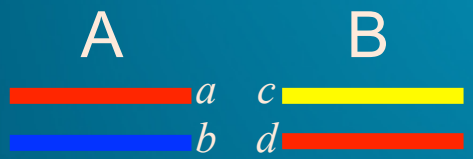
- Situation is very trivial for the case of **A**
- For other matrices, the aim is thus to determine a set of $\mathbf{b}_j$ with as much advantages as in **A**:
  - ⇔easy to determine
  - ⇔involving few computations
  - ⇔sparse
  - ⇔having linear computation cost with size of matrix

Application to the case of the inverse of G

(Faux, Gengler and Misztal, 2012)

The genomic relationship matrix

- A few words about genomic relationship...
  - Observed (or, at least, sampled) vs. expected
  - Some introduction with Bömcke and Gengler (2009)



$$fM_{x,y;l} = \frac{1}{4}\left(S_{ac} + S_{ad} + S_{bc} + S_{bd}\right) \quad ; \quad TA_{x,y;l} = 2\,fM$$

$$TA_{x,y} = \frac{\sum_{l=1}^{m} TA_{x,y;l}}{m} \quad ; \quad \mathbf{TA} = \begin{bmatrix} TA_{1,1} & \cdots & TA_{1,1} \\ \vdots & \ddots & \vdots \\ TA_{1,1} & \cdots & TA_{1,1} \end{bmatrix}$$

The genomic relationship matrix

- When working with a SNP matrix (**M**; *n* x *m*):

$$\mathbf{M}=\begin{bmatrix} 1 & 2 & 1 & 1 & 0 & \dots \\ 0 & 0 & 1 & 1 & 0 & \dots \\ 1 & 2 & 0 & 1 & 1 & \dots \end{bmatrix}; \quad \mathbf{Z}=\mathbf{M}-\mathbf{1}$$

$$\mathbf{TA}=\frac{\mathbf{ZZ}'+m}{m}$$

- Van Raden's **G**:

$$\mathbf{f}=\begin{bmatrix} MAF_1 & \dots & MAF_m \end{bmatrix} \; ; \quad d=2\cdot\mathbf{f}'\cdot(\mathbf{1}-\mathbf{f})$$

$$\mathbf{P}=\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\otimes\begin{bmatrix} 2f_1-1 & \dots & 2f_m-1 \end{bmatrix} \; ; \quad \mathbf{Z}_g=\mathbf{M}-\mathbf{P} \; ; \quad \mathbf{G}=\frac{\mathbf{ZZ}'}{d}$$

Approximation of the inverse of the factorization of G

- Purpose: decrease computation time
- Solution: create the $\mathbf{b}_j$ using **OLS on close-related**
- Why? Go back to tabular method...

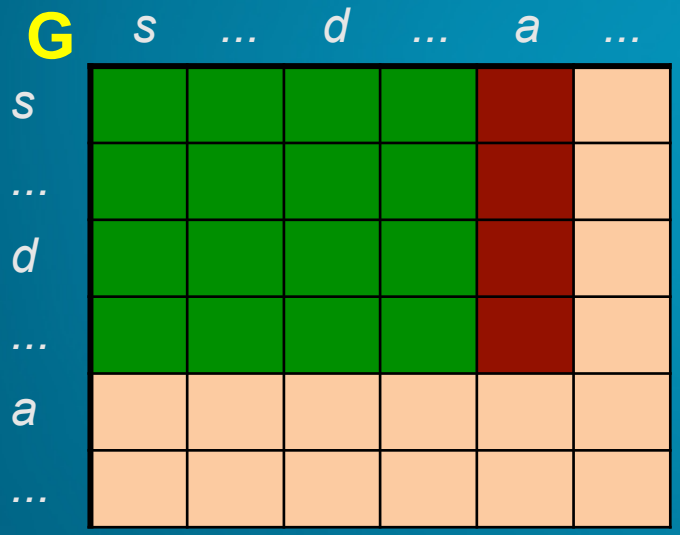
... Model used in tabular method is back in the inverse decomposition!

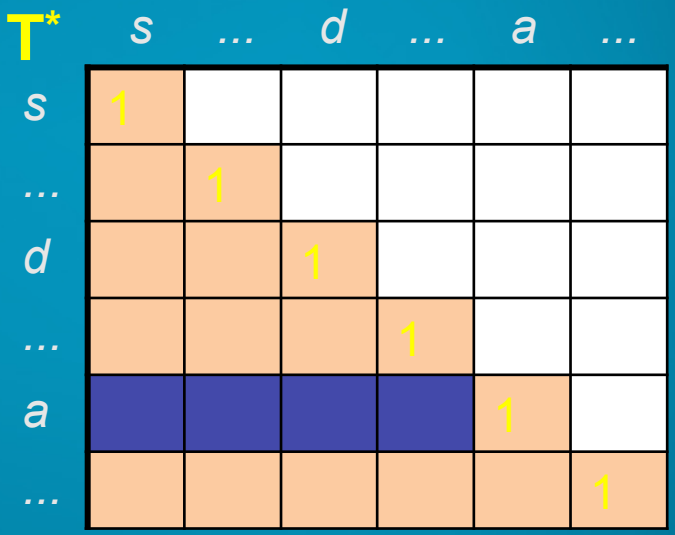
- Why? It ensures **D** to be diagonal:

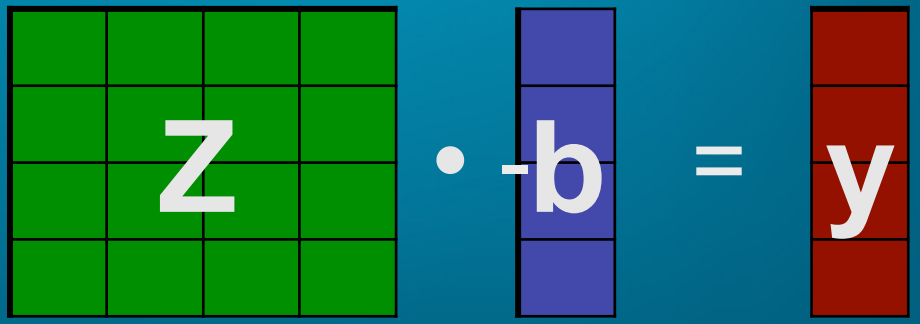
$$\mathbf{A}^{-1} = (\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1} \quad \Rightarrow \quad \mathbf{D} = \mathbf{T}^{-1}\mathbf{A}(\mathbf{T}^{-1})'$$

Approximation of the inverse of the factorization of G

- Solutions are computed by OLS using a simple model:







Approximation of the inverse of the factorization of G

- Solutions are computed by OLS using a simple model:

$$\mathbf{b} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

$$\mathbf{b} = (\mathbf{Z})^{-1}(\mathbf{Z})^{-1}(\mathbf{Z})\mathbf{y}$$

$$\Rightarrow \mathbf{b} = (\mathbf{Z})^{-1}\mathbf{y}$$

Approximation of the inverse of the factorization of G

- Solutions are computed by OLS using a simple model:

$$\mathbf{b} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

$$\mathbf{b} = (\mathbf{Z})^{-1}(\mathbf{Z})^{-1}(\mathbf{Z})\mathbf{y}$$

$$\Rightarrow \mathbf{b} = (\mathbf{Z})^{-1}\mathbf{y}$$

- Restriction to the close-family (Ω) of an animal
  - *p* is a « genomic » threshold that defines close-family

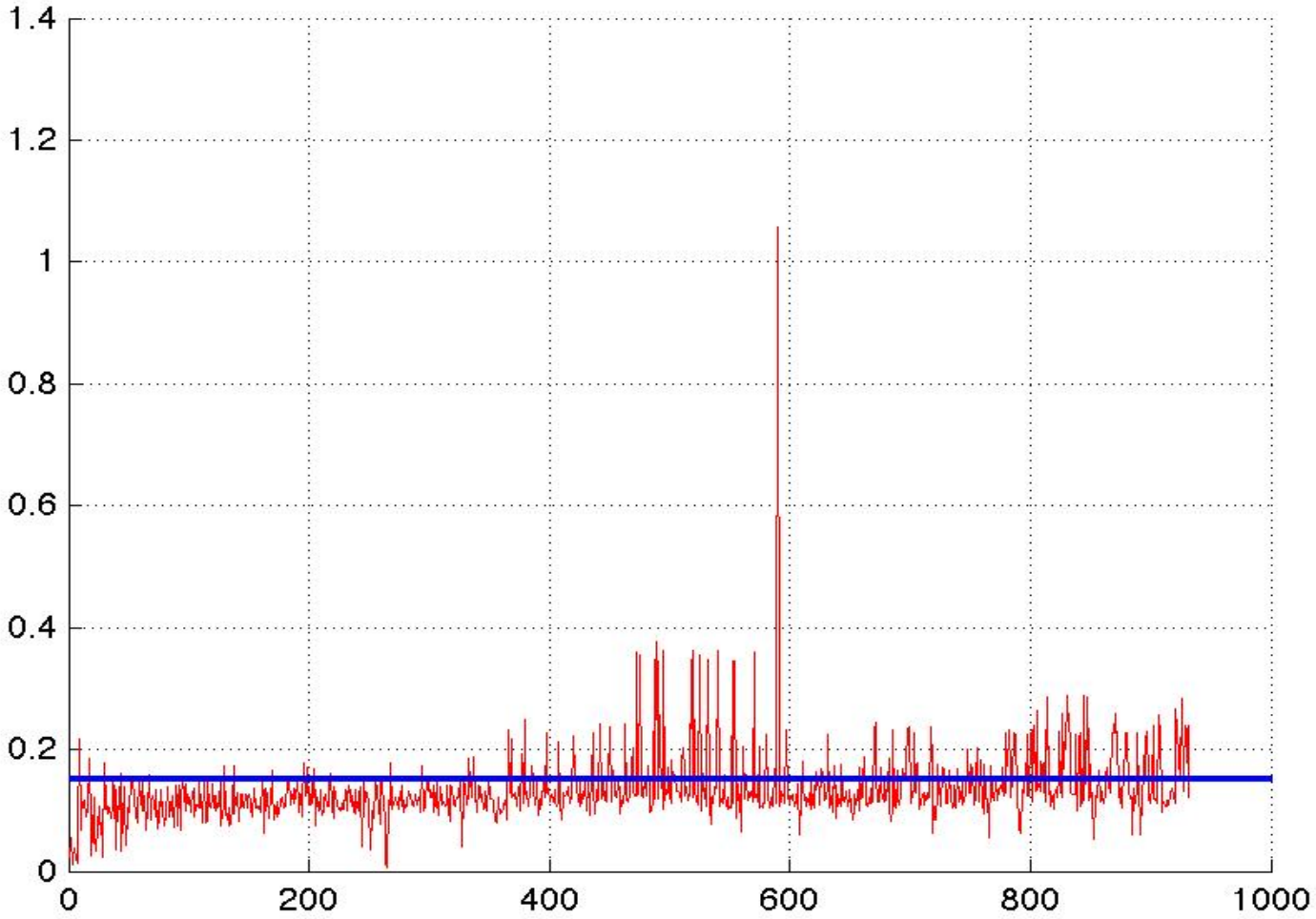
$$\Omega_i = \{k : k < i, g_{ik} \geq p\}$$

Computation of D^{-1} and G^{-1}

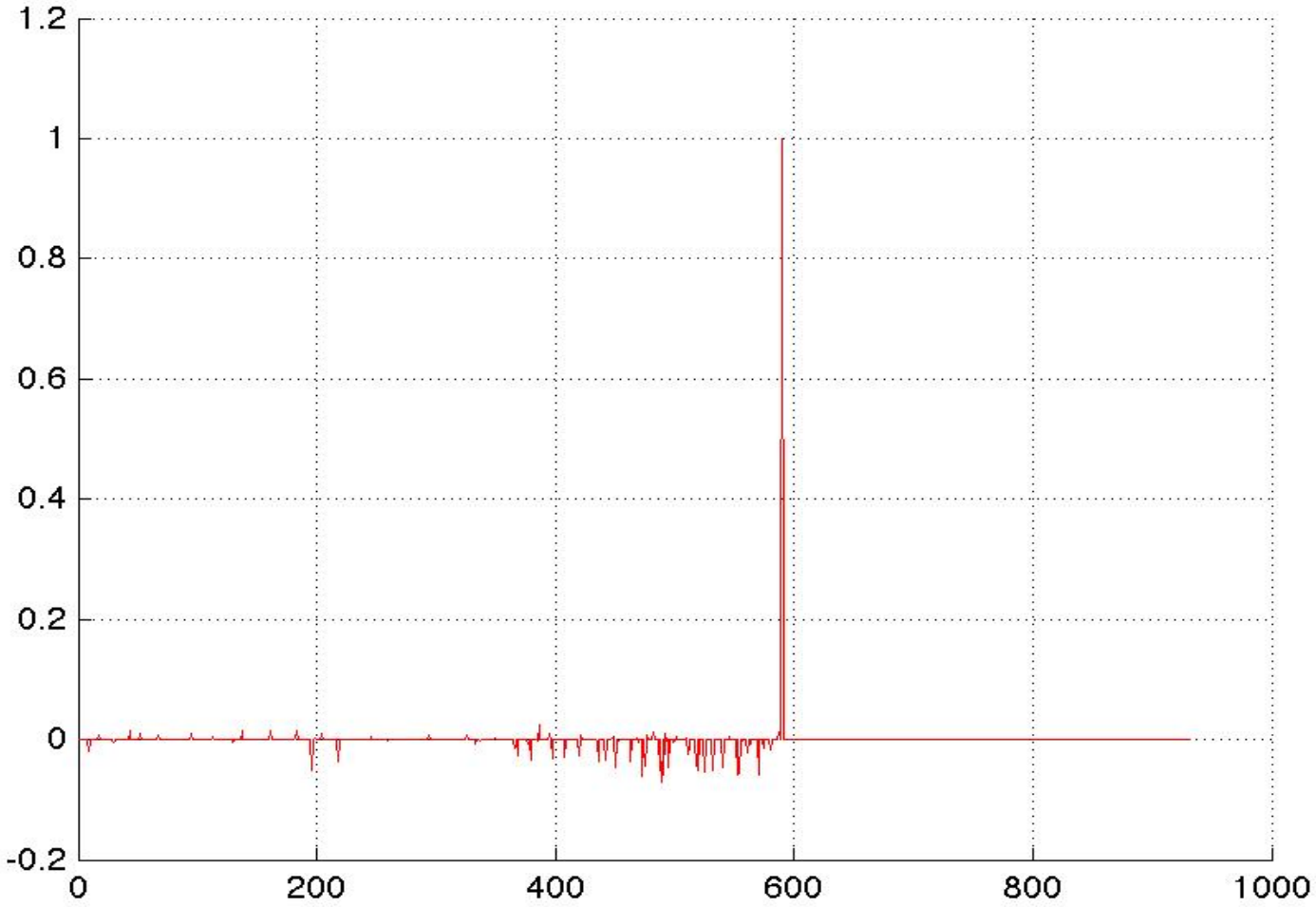
- Use of the “backward” equation to get **D**:

$$\mathbf{D} = \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})'$$

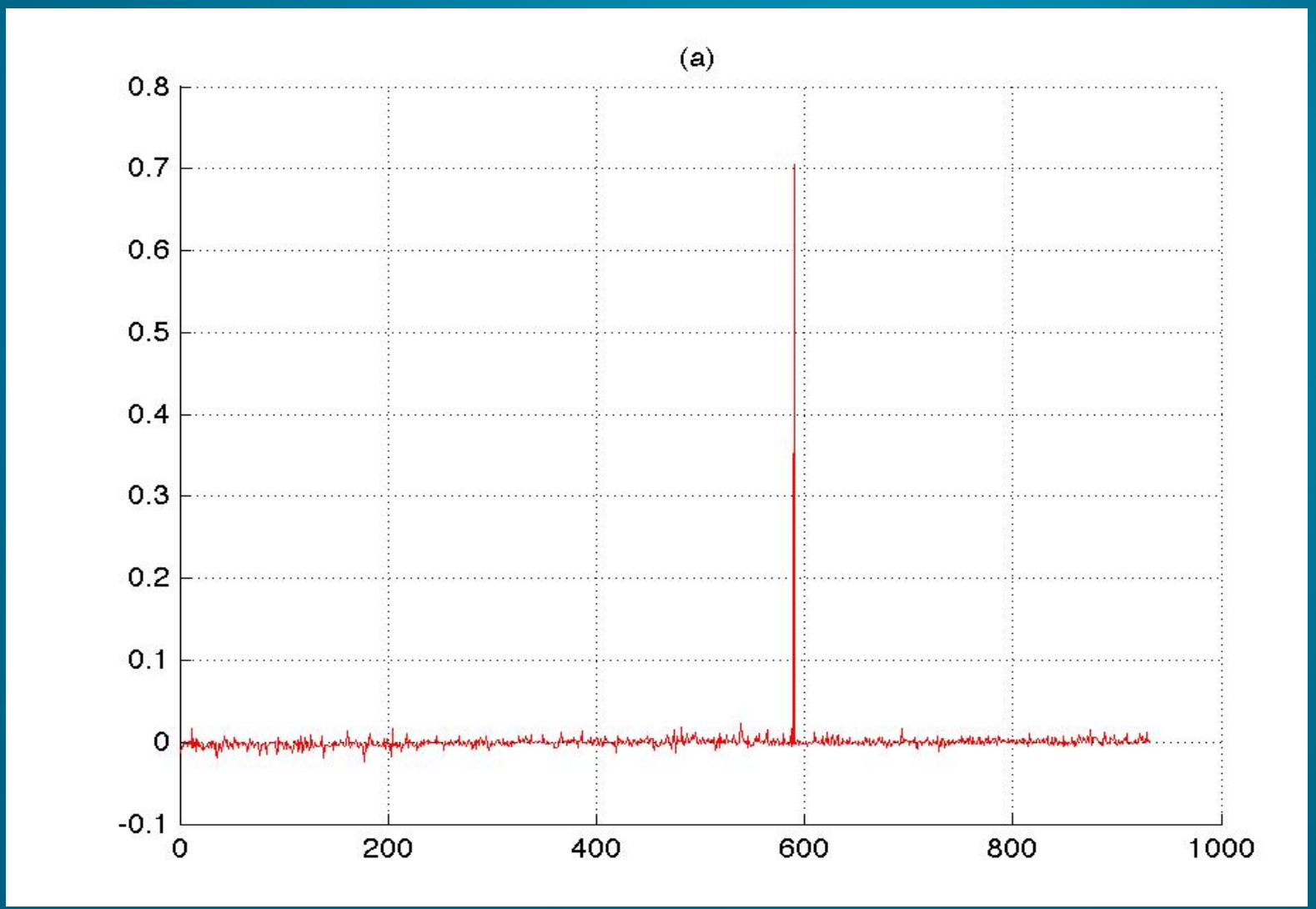
Computation of D^{-1} and G^{-1}



Computation of D^{-1} and G^{-1}



Computation of D^{-1} and G^{-1}



Computation of D^{-1} and G^{-1}

- Use of the "backward" equation to get **D**:

$$\mathbf{D} = \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})'$$

- Next? 2 options:
  1) Take a diagonal element of **D** and invert it
  2) Recursively process the remaining **D** as **G**

- Final equation:

$$\mathbf{G}^{-1} = (\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1}$$

Acknowledgements

- Dr. G. Gorjanc (UL), Dr. F. Colinet and Mr. J. Vandenplas (ULg)
- Wallonie-Bruxelles International (WBI)
- Government of Republic of Slovenia

- Fonds National de la Recherche Luxembourg (FNR)
- **Host institution:** Gembloux Agro-Bio Tech - University of Liège
- **Collaborating institutions**:
  - CONVIS s.c. (Luxembourg)
  - UGA, Animal and Dairy Science Department (USA)